Easily Verifiable Conditions for the Convergence of the Markov Chain Monte Carlo Method

Jayaram Sethuraman*
Florida State University
Tallahassee, FL 32306

and

Indian Statistical Institute
Bangalore 560 059

December 1995

FSU Techincal Report Number **M 906**
USARO Technical Report Number **D-136**

*Research supported by Army Research Office Grant DAAH04-93-G-0201

Easily Verifiable Conditions for the Convergence of the Markov Chain Monte Carlo Method

Jayaram Sethuraman*
Florida State University
Tallahassee, FL 32306

and

Indian Statistical Institute
Bangalore 560 059

December 1995

Abstract

The Markov Chain Chain Monte Carlo (MCMC) method, which is a special case of the Gibbs sampler, is a very powerful method to simulate from complicated distributions arising in many contexts, including image analysis, computational Bayesian analysis, and so on. Existing results that ensure that this method will converge involve conditions which are difficult to verify in practice, and most practitioners, convinced that their particular problem will not be pathological and give up verifying altogether. This paper gives a new set of sufficient conditions which are easy to verify in most applications.

Key words and phrases: Markov Chains, Gibbs Sampling, Computational Bayes Methods, Ergodic Theorem.

Abbreviated title: Markov Chain Monte Carlo Methods.

AMS (1991) subject classifications: Primary 60J05; secondary 65U05, 60B10.

*Research supported by Army Research Office Grant DAAH04-93-G-0201

1 The First Example

We begin with a familiar example found in the book Rao (1965) "Linear Statistical Inference and its Applications" on data on blood groups in human populations use it to illustrate Markov Chain Monte Carlo methods. Every human being can be classified into one of four blood groups O, A, B and AB. The inheritance of these blood groups is controlled by three allelomorphic genes O, A and B, where O is recessive to both A and B. If r, p and q are the gene frequencies of O, A and B, then the probabilities of of the six genotypes and the four phenotypes, under random mating, and a typical data on a human population of size N can be represented by the following table:

Group		Probabilities		Frequency	
Phenotype	Genotype	Phenotype	Genotype	Observed	Unobserved
O	OO	r^2	r^2	$n(O)$	
A	$\begin{cases} AA \\ AO \end{cases}$	p^2+2pr	$\begin{cases} p^2 \\ 2pr \end{cases}$	$n(A)$	$\begin{cases} n(AA) \\ n(A)-n(AA) \end{cases}$
B	$\begin{cases} BB \\ BO \end{cases}$	q^2+2qr	$\begin{cases} q^2 \\ 2qr \end{cases}$	$n(B)$	$\begin{cases} n(BB) \\ n(B)-n(BB) \end{cases}$
AB	AB	$2pq$	$2pq$	$n(AB)$	
Totals		1	1	N	

Here $n(O), n(A), n(B)$ and $n(AB)$, which will be called the *data*, are the observed frequencies of the four blood groups in a population of size N. The frequencies $n(AA)$ and $n(BB)$ of the genotypes AA and BB cannot be observed. The problem is to estimate the probabilities p, q and r.

The *data* follow a simple multinomial distribution with 4 cells, where the cell probabilities are functions of the parameters of interest, and the likelihood is proportional to

$$r^{2n(O)}(p^2+2pr)^{n(A)}(q^2+2qr)^{n(B)}(2pq)^{n(AB)}.$$

The maximum likelihood equations are not easy to solve directly and Rao ((1965) pp. 305–309) suggests the standard method of scoring to obtain the maximum likelihood estimates. How will a Bayesian approach this problem? Since $p+q+r=1$, it is natural to put $\mathcal{D}(\alpha_1, \alpha_2, \alpha_3)$, the Dirichlet distribution with parameters $\alpha_1 > 0, \alpha_2 > 0, \alpha_3 > 0$, as a prior distribution for (p, q, r). The next step is to obtain the posterior distribution conditional on the *data*. Once again, this turns out to be an untractable problem. However, if the unobserved frequencies $n(AA)$ and $n(BB)$ were available, then the posterior of (p, q, r) given the $(data, n(AA), n(BB))$ is easy to obtain. Note that the likelihood of the $data, n(AA)$

and $n(BB)$ again comes from a multinomial distribution with 6 cells and is proportional to

$$p^{n(AA)+n(A)+n(AB)}q^{n(BB)+n(B)+n(AB)}r^{2n(O)+n(A)-n(AA)+n(B)-n(BB)}.$$

Denote (p,q,r) by $\mathbf{Y} = (\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \mathbf{Y}^{(3)})$ and $(n(AA), n(BB))$ by $\mathbf{Z} = (\mathbf{Z_1}, \mathbf{Z_2})$. From the above remarks, the conditional distribution of $\mathbf{Y}$ given $(data, \mathbf{Z})$ can be written as

$$\mathcal{L}\{\mathbf{Y}|data, \mathbf{Z})\} = \mathcal{D}(\alpha_1', \alpha_2', \alpha_3') \tag{1.1}$$

where

$$\begin{aligned}
\alpha_1' &= \alpha_1 + n(AA) + n(A) + n(AB), \\
\alpha_2' &= \alpha_2 + n(BB) + n(B) + n(AB) \quad \text{and} \\
\alpha_3' &= \alpha_3 + 2n(O) + n(A) - n(AA) + n(B) - n(BB)).
\end{aligned}$$

It is easy to write down the conditional distribution of the unobserved frequencies $\mathbf{Z}$ given the $data, \mathbf{Y}$ as

$$\mathcal{L}\{\mathbf{Z}|data, \mathbf{Y}\} = B(n(A), \frac{p}{p+2r}) \times B(n(B), \frac{q}{q+2r}) \tag{1.2}$$

where $B(M, \theta)$ stands for the Binomial distribution with M trials and probability of success θ.

Notice that equations (1.1) and (1.2) give us the conditional distributions $\mathcal{L}\{\mathbf{Y}|data, \mathbf{Z}\}$ and $\mathcal{L}\{\mathbf{Z}|data, \mathbf{Y}\}$. The Gibbs sampler, which is a special case of the Markov Chain Monte Carlo method, can be used to obtain a Markov chain $\mathbf{X_0} = (\mathbf{Y_0}, \mathbf{Z_0}), \mathbf{X_1} = (\mathbf{Y_1}, \mathbf{Z_1}), \ldots$ such that the distribution of $\mathbf{X}_n = (\mathbf{Y}_n, \mathbf{Z}_n)$ will converge to $\mathcal{L}\{(\mathbf{Y}, \mathbf{Z})|data\}$ as $n \to \infty$. By considering just the marginals $\mathbf{Y_0}, \mathbf{Y_1}, \ldots$ we see that $\mathbf{Y}_n$ converges in distribution to required posterior distribution $\mathcal{L}\{(p,q,r)|data\} = \mathcal{L}\{\mathbf{Y}|data\}$. We could also use other methods based on Markov chain theory to obtain better approximations to $\mathcal{L}\{\mathbf{Y}|data\}$. Finally, we can approximate $E(\mathbf{Y}|data)$ which is the Bayes estimate of the vector (p,q,r). This would be the Bayesian answer to the method of scoring for maximum likelihood estimates.

How is the Markov chain $\mathbf{X_0}, \mathbf{X_1}, \ldots$ generated? Fix arbitrary values for $(\mathbf{Y_0}, \mathbf{Z_0})$. For $n = 0, 1, \ldots$ generate $\mathbf{Y_{n+1}}$ from the distribution $\mathcal{L}\{\mathbf{Y}|data, \mathbf{Z}_n\}$ as given in (1.1) and generate $\mathbf{Z}_{n+1}$ from the distribution $\mathcal{L}\{\mathbf{Z}|data, \mathbf{Y}_{n+1}\}$ as given in (1.2). This way we generate $(\mathbf{Y_1}, \mathbf{Z_1}), (\mathbf{Y_2}, \mathbf{Z_2}), \ldots$. It is easy to see that this is a Markov chain whose transition function can be expressed in terms of $\mathcal{L}\{\mathbf{Y}|data, \mathbf{Z}\}$ and $\mathcal{L}\{\mathbf{Z}|data, \mathbf{Y}\}$ and that $\mathcal{L}\{(\mathbf{Y}, \mathbf{Z})|data\}$ is an invariant distribution for this transition function.

2 The Second Example

Let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ be i.i.d random variables with unknown distribution P. Let $C_1, C_2, \ldots, C_n$ be subsets of the real line, some of which may be singleton sets. Suppose that the *data* gives the information $Y_i \in C_i, i = 1 \ldots, n$. If the $C_i = \{c_i\}, i = 1, \ldots, n$ are all singletons then we have observed the actual values of $Y_1, \ldots, Y_n$. If the some C_i's are singletons and other C_i's are sets of the form $[c_i, \infty)$, then this corresponds to case

of right censoring. The frequentist solution to estimating P is the usual Kaplan-Meyer estimate. What if one were a Bayesian, and one uses a Dirichlet prior $\mathcal{D}_\alpha$, where α is a finite measure on the real line? Suppose that there are m uncensored observations and $n-m$ censored observations, that is without loss of generality that $C_1 = \{c_1\}, \ldots, C_m = \{c_m\}$ are singletons and the remaining C_i's are not singletons. Then, the *data* gives us the information that $Y_1 = c_1, \ldots, Y_m = c_m$ and that $Y_{m+1} \in C_{m+1}, \ldots, Y_n \in C_n$. Let $\mathbf{V} = (Y_{m+1}, \ldots, Y_n)$ be the actual unobserved values of the censored observations. What is the posterior distribution of P given *data*? From the standard theory of Dirichlet distributions, see for instance Ferguson (1972) or Sethuraman (1994), the posterior distribution of P given $Y_1 = c_1, \ldots, Y_m = c_m$ is $\mathcal{D}_{\alpha'}$ where $\alpha' = \alpha + \sum_{i=1}^{m} \delta_{c_i}$. As before, the posterior distribution of P given *data* is not tractable. For any probability measure μ and set B with $\mu(B) > 0$, let $\mu_B(C) = \frac{\mu(C \cap B)}{\mu(B)}$ be the restriction to B. Then we have the following two facts:

$$\text{the conditional distribution of } P \text{ given } \{data, \mathbf{V}\} \text{ is } \mathcal{D}_\beta \text{ where } \beta = \alpha' + \sum_{i=m+1}^{n} \delta_{\mathbf{V}_i}, \quad (2.1)$$

and

$$\text{the conditional distribution of } \mathbf{V} \text{ given } \{data, P\} \text{ is } \prod_{i=m+1}^{n} P_{C_i}. \quad (2.2)$$

Starting from arbitrary values $(P_0, \mathbf{V}_0)$, we can carry out the Markov Chain Monte Carlo Method to generate a Markov Chain $(P_1, \mathbf{V}_1), (P_2, \mathbf{V}_2), \ldots$. We can hope that its distribution will converge to the joint distribution of $(P, \mathbf{V})$ given $\mathbf{Z}$, and the required posterior distribution is obtained from here by taking the marginal distribution of P. A crucial intermediate step in this Markov Chain Monte Carlo follows from the constructive definition given is Sethuraman (1994). See Doss (1996) for details. Once again, the question arises whether this Markov Chain will converge to the desired conditional distribution.

3 The Markov Chain Monte Carlo Method

Examples such as the one described in Sections 1 and 2 arise in many areas of Statistics. In each of these problems there is a probability distribution π on a measurable space $(\mathcal{X}, \mathcal{B})$, and we are interested in estimating characteristics of it such as $\pi(E)$ or $\int f d\pi$ where $E \in \mathcal{B}$ and f is a bounded measurable function. Even when π is fully specified one may have to resort to methods like Monte Carlo simulation, especially when π is not computationally tractable. For this one uses the available huge literature on generation of random variables from an explicitly or implicitly described probability distribution π. Generally these methods require $\mathcal{X}$ to be the real line or require that π have special features, such as a structure in terms of independent real valued random variables. When one cannot generate random variables with distribution π one has to be satisfied with looking for a sequence of random variables $X_1, X_2, \ldots$ whose distributions converge to π and using X_n with a large index n as an observation from π. This is called the Markov Chain Monte Carlo Method. The preceding discussion of blood group data from Dr. C. R. Rao's book is an illustrative example. In this example, π is the posterior distribution $\mathcal{L}\{\mathbf{Y}, \mathbf{Z} | data\}$ and an example of a functionals of interest may be $E(\mathbf{Y}|data)$ the Bayes estimate of $\mathbf{Y}$.

Let P be a transition probability function on a measurable space $(\mathcal{X}, \mathcal{B})$, i.e. P is a function on $\mathcal{X} \times \mathcal{B}$ such that for each $x \in \mathcal{X}$, $P(x, \cdot)$ is a probability measure on $(\mathcal{X}, \mathcal{B})$, and for each $C \in \mathcal{B}$, $P(\cdot, C)$ is a measurable function on $(\mathcal{X}, \mathcal{B})$. Suppose that π is a probability measure on $(\mathcal{X}, \mathcal{B})$ which is invariant for the Markov chain, i.e.

$$\pi(C) = \int P(x, C)\pi(dx) \quad \text{for all } C \in \mathcal{B}. \tag{3.1}$$

We fix a starting point x_0, generate an observation X_1 from $P(x_0, \cdot)$, generate an observation X_2 from $P(X_1, \cdot)$, etc. This generates the Markov chain $x_0 = X_0, X_1, X_2, \ldots$. In order to make use of the Markov chain $\{X_n\}_{n=0}^{\infty}$ to get some information about π, one needs results of the form:

(a) *Ergodicity:* For all or for "most" starting values x, the distribution of X_n converges to π in a suitable sense, for example

(a1) *Variation norm ergodicity:* $\sup_{C \in \mathcal{B}} |P^n(x, C) - \pi(C)| \to 0$, or

(a2) *Variation norm mean ergodicity:* $\sup_{C \in \mathcal{B}} |\frac{1}{n} \sum_{j=1}^{n} P^j(x, C) - \pi(C)| \to 0$.

(b) *Law of large numbers:* For all or for most starting values x, for each $C \in \mathcal{B}$,

$$\frac{1}{n} \sum_{j=1}^{n} I_C(X_j) \to \pi(C) \quad \text{for a.e. realization of the chain,}$$

and for each f with $\int |f| d\pi < \infty$,

$$\frac{1}{n} \sum_{j=1}^{n} f(X_j) \to \int f d\pi \quad \text{for a.e. realization of the chain.}$$

Then, we may estimate π for example by generating G such chains in parallel, obtaining independent observations $X_n^{(1)}, \ldots, X_n^{(G)}$, or by running one (or a few) very long chains.

4 Main Results

Our goal is to find conditions on a given Markov chain or rather on its transition function $P(\cdot, \cdot)$ so that some or all of the conditions (a) and (b) above hold, assuming that P admits an invariant probability measure π. In Markov Chain Monte Carlo applications, the probability measure π of interest is *by construction* the invariant probability measure of the Markov chain.

When $\{X_n\}$ is a Markov chain with a countable state space, say $\{1, 2, \ldots\}$, and transition probability matrix $P = (p_{i,j})$, the existence of an invariant probability distribution π and the *irreducibility condition* that there exists a state i_0 such that from any initial state i, there is positive probability that the chain eventually hits i_0, are enough to guarantee that (i) the chain $\{X_n\}$ is recurrent in an appropriate sense, (ii) conditions (b) and (a2) above hold, and (iii) when an additional aperiodicity condition also holds, then (a1) above also holds. These facts are well known; see for instance, Hoel, Port and Stone (1972).

A natural question is whether this is true for general state space Markov chains. In particular, when (3.1) holds, is there a form of the irreducibility condition under which some or all of (a) and (b) above hold?

The Markov chain literature has a number of results in this direction; see Orey (1971), Athreya and Ney (1978) and Nummelin (1984). Under a condition known as Harris recurrence (see below) the existence of an invariant distribution π implies mean ergodicity (condition (a2)) and the laws of large numbers (condition (b)). Unfortunately, Harris recurrence is not an easy condition to verify in general, and it is much stronger than irreducibility.

The main point of this paper is to show that when (3.1) holds, a simple irreducibility condition ((4.3) below) is enough to yield (a2) and (b). An additional aperiodicity condition yields (a1) as well. This provides a complete generalization of the results for the countable case. *It is worth noting that recurrence emerges as a consequence of (3.1) and the irreducibility condition (4.3), and is not imposed as a hypothesis.*

Before stating our main theorems, we will need a few definitions. For any set $C \in \mathcal{B}$, let $N_n(C) = \sum_{m=1}^{n} I(X_m \in C)$ and $N(C) = \sum_{m=1}^{\infty} I(X_m \in C)$ be the number of visits to C by time n and the total number of visits to C, respectively. The expectations of $N_n(C)$ and $N(C)$, when the chain starts at x, are given by $G_n(x, C) = \sum_{m=1}^{n} P^m(x, C)$ and $G(x, C) = \sum_{m=1}^{\infty} P^m(x, C)$, respectively. Define $T(C) = \inf\{n : n > 0,\ X_n \in C\}$ to be the first time the chain hits C, after time 0. Note that $P_x(T(C) < \infty) > 0$ is equivalent to $G(x, C) > 0$.

The set $A \in \mathcal{B}$ is said to be *accessible from* x if $P_x(T(A) < \infty) > 0$. Let ρ be a probability measure on $(\mathcal{X}, \mathcal{B})$. The Markov chain is said to be *ρ-recurrent* (or Harris recurrent with respect to ρ) if for every A with $\rho(A) > 0$, $P_x(T(A) < \infty) = 1$ for all $x \in \mathcal{X}$. The chain is said to be *ρ-irreducible* if every set A with $\rho(A) > 0$ is accessible from all $x \in \mathcal{X}$. The set A is said to be *recurrent* if $P_x(T(A) < \infty) = 1$ for all $x \in \mathcal{X}$.

For the case where the σ-field $\mathcal{B}$ is separable, there is a very useful equivalent definition of ρ-irreducibility of a Markov chain. In this case, we can deduce from Theorem 2.1 of Orey (1971), on the existence of "C-sets," that ρ-irreducibility of a Markov chain implies that there exist a set $A \in \mathcal{B}$ with $\rho(A) > 0$, an integer n_0, and a number $\epsilon > 0$ satisfying

$$P_x(T(A) < \infty) > 0 \quad \text{for all } x \in \mathcal{X}, \tag{4.1}$$

and

$$x \in A,\ C \in \mathcal{B} \quad \text{imply} \quad P^{n_0}(x, C) \geq \epsilon \rho(C \cap A). \tag{4.2}$$

Let $\rho_A(C) = \frac{\rho(C \cap A)}{\rho(A)}$. This is well defined because $\rho(A) > 0$. The set function ρ_A is a probability measure satisfying $\rho_A(A) = 1$. Note that (4.1) simply states that A is accessible from all $x \in \mathcal{X}$ and this condition does not make reference to the probability measure ρ. Condition (4.2) states that uniformly in $x \in A$, the n_0-step transition probabilities from x into subsets of A are bounded below by ϵ times ρ. That (4.1) and (4.2) imply ρ_A-irreducibility is, of course, immediate. This alternative definition of ρ_A-irreducibility, which applies to nonseparable σ-fields as well, will be usually much easier to verify in Markov chain simulation problems. By replacing ρ by ρ_A, we can also assume with no loss of generality that ρ is a probability measure with $\rho(A) = 1$ when verifying Condition (4.2).

We denote the greatest common divisor of any subset $\mathcal{M}$ of integers by g.c.d.$(\mathcal{M})$.

We now state two theorems which hold for general Markov chains. They give sufficient conditions for the Markov Chain Monte Carlo method to be successful and constitute the

main results of this paper. The proofs of these theorems can be found in Athreya, Doss and Sethuraman (1996).

Theorem 1 *Suppose that the Markov chain $\{X_n\}$ with transition function $P(x, C)$ has an invariant probability measure π, i.e. (3.1) holds. Suppose that there is a set $A \in \mathcal{B}$, a probability measure ρ with $\rho(A) = 1$, a constant $\epsilon > 0$, and an integer $n_0 \geq 1$ such that*

$$\pi\Big\{x : P_x(T(A) < \infty) > 0\Big\} = 1, \tag{4.3}$$

and

$$P^{n_0}(x, \cdot) \geq \epsilon\rho(\cdot) \quad \text{for each } x \in A. \tag{4.4}$$

Suppose further that

$$g.c.d.\Big\{m : \text{there is an } \epsilon_m > 0 \text{ such that } P^m(x, \cdot) \geq \epsilon_m\rho(\cdot) \text{ for each } x \in A\Big\} = 1. \tag{4.5}$$

Then there is a set D such that

$$\pi(D) = 1 \quad \text{and} \quad \sup_{C\in\mathcal{B}} |P^n(x, C) - \pi(C)| \to 0 \quad \text{for each } x \in D. \tag{4.6}$$

Theorem 2 *Suppose that the Markov chain $\{X_n\}$ with transition function $P(x, C)$ satisfies Conditions (3.1), (4.3) and (4.4). Then*

$$\sup_{C\in\mathcal{B}} \Big|\frac{1}{n_0}\sum_{r=0}^{n_0-1} P^{mn_0+r}(x, C) - \pi(C)\Big| \to 0 \quad \text{as } m \to \infty \quad \text{for } [\pi]\text{-almost all } x, \tag{4.7}$$

and hence

$$\sup_{C\in\mathcal{B}} \Big|\frac{1}{n}\sum_{j=1}^{n} P^j(x, C) - \pi(C)\Big| \to 0 \quad \text{as } n \to \infty \quad \text{for } [\pi]\text{-almost all } x. \tag{4.8}$$

Let $f(x)$ be a measurable function on $(\mathcal{X}, \mathcal{B})$ such that $\int \pi(dy)|f(y)| < \infty$. Then

$$P_x\Big\{\frac{1}{n}\sum_{j=1}^{n} f(X_j) \to \int \pi(dy)f(y)\Big\} = 1 \quad \text{for } [\pi]\text{-almost all } x \tag{4.9}$$

and

$$\frac{1}{n}\sum_{j=1}^{n} E_x(f(X_j)) \to \int \pi(dy)f(y) \quad \text{for } [\pi]\text{-almost all } x. \tag{4.10}$$

Variants of these theorems form a main core of interest in the Markov chain literature. However, most of this literature makes strong assumptions such as the existence of a recurrent set A and proves the *existence* of an invariant probability measure before establishing (4.6) and (4.7). Theorems 1 and 2 exploit the existence of an invariant probability measure, which is given to us "for free" in the Markov chain simulation context, and establish the ergodicity or mean ergodicity under minimal and easily verifiable assumptions. For example, we have already noted that in the context of the Markov chain simulation method, we really need to check only (4.3), (4.4), and (4.5). To show (4.3) in most cases one will establish that $P_x(T(A) < \infty) > 0$ for all x. Condition (4.5) is usually called the *aperiodicity* condition and is automatically satisfied if (4.4) holds with $n_0 = 1$. Condition (4.4) holds if for each $x \in A$, $P^{n_0}(x, \cdot)$ has a non-trivial absolutely continuous component with respect to some measure ρ and the associated density $p^{n_0}(x, y)$ satisfies $\inf_{x,y\in A} p^{n_0}(x, y) > 0$ for some A with $\rho(A) > 0$.

5 Back to the Illustrative Examples

Consider the special case where $\mathcal{X} = \times_{j=1}^{2} \mathcal{X}^{(j)}$ and $\mathcal{B} = \times_{j=1}^{2} \mathcal{B}^{(j)}$. We give below two theorems that give easily verifiable conditions for the conclusions of Theorems 1 and 2 to hold. We will use one of these theorems to show that the conditions of Theorems 1 and 2 hold in the example of Section 1 on data on blood groups of humans.

Consider the Markov Chain Monte Carlo algorithm for generating observations from the joint distribution π of $(X^{(1)}, X^{(2)})$ as described in Section 3.

Theorem 3 *Suppose that the conditional distributions $\pi_{X^{(1)}|X^{(2)}=x^{(2)}}$ and $\pi_{X^{(2)}|X^{(1)}=x^{(1)}}$ have densities, say $p_{X^{(1)}|X^{(2)}}(x^{(1)}|x^{(2)})$ and $p_{X^{(2)}|X^{(1)}}(x^{(2)}|x^{(1)})$, respectively with respect to some dominating measures $\rho^{(1)}$ and $\rho^{(2)}$. Suppose further that for each $i = 1, 2$ there is a set $A^{(i)}$ with $\rho^{(i)}(A^{(i)}) > 0$, and a $\delta > 0$ such that*

$$p_{X^{(1)}|X^{(2)}}(x^{(1)}|x^{(2)}) > 0 \tag{5.1}$$

whenever

$$x^{(2)} \in A^{(2)} \quad \text{and} \quad x^{(1)} \text{ is arbitrary,}$$

and

$$p_{X^{(1)}|X^{(2)}}(x^{(1)}|x^{(2)}) > \delta \quad \text{and} \quad p_{X^{(2)}|X^{(1)}}(x^{(2)}|x^{(1)}) > \delta \quad \text{whenever} \quad x^{(1)} \in A^{(1)}, x^{(2)} \in A^{(2)}. \tag{5.2}$$

Then Conditions (4.3) and (4.4) are satisfied with $n_0 = 1$. Thus, (4.5) is also satisfied, and the conclusions of Theorems 1 and 2 hold.

Theorem 4 *Suppose that the conditional distributions $\pi_{X^{(2)}|X^{(1)}=x^{(1)}}$ has a density, say $p_{X^{(2)}|X^{(1)}}(x^{(2)}|x^{(1)})$ with respect to some dominating measure $\rho^{(2)}$. Suppose that there are sets $A^{(1)}$ and $A^{(2)}$, and a $\delta > 0$ such that*

$$\pi_{X^{(1)}|X^{(2)}}(A^{(1)}|x^{(2)}) > 0 \tag{5.3}$$

for all $x^{(2)}$,

$$\pi_{X^{(1)}|X^{(2)}}(A^{(1)}|x^{(2)}) > \delta \tag{5.4}$$

for all $x^{(2)} \in A^{(2)}$, and

$$p_{X^{(2)}|X^{(1)}}(x^{(2)}|x^{((1))}) > \delta \quad \text{whenever} \quad x^{(1)} \in A^{(1)}, x^{(2)} \in A^{(2)}. \tag{5.5}$$

Then conditions (4.3) and (4.4) are satisfied with $n_0 = 1$. Thus, (4.5) is also satisfied, and the conclusions of Theorems 1 and 2 hold.

We will verify the conditions of Theorem 3 in the example of blood group data from humans in Section 1. Here $\mathcal{X}^{(1)} = R^3, \mathcal{X}^{(2)} = \{0, \ldots, n(A)\} \times \{0, \ldots, n(B)\}, X^{(1)} = \mathbf{Y}$ and $X^{(2)} = \mathbf{Z}$. We can take $\rho^{(1)}$ as the Lebesgue measure of R^3 and $\rho^{(2)}$ as the counting measure on $\mathcal{X}^{(2)}$. Put $A^{(1)} = [0.23, 0.43] \times [0.23, 0.43] \times [0.23, .43]$ and $A^{(2)} = \mathcal{X}^{(2)}$. It is easy to verify that $\rho^{(i)}(A^{(i)}) > 0$ for $i = 1, 2$ and that conditions (5.1) and (5.2) are satisfied for some $\delta > 0$, since the Dirichlet distribution has a positive density function, $A^{(2)}$ is a finite

set and the Binomial distribution with parameter in $[0.23, 0.43]$ has a positive frequency function.

The verification of the conditions of Theorems 1 and 2 for the Bayesian solution to the Kaplan-Meyer problem is not quite straightforward since the state space of $(P, \mathbf{V})$ is more complicated. We therefore consider only the Markov chain $\mathbf{V}_0, \mathbf{V}_1, \mathbf{V}_2, \ldots$ whose state space is R^{n-m}. By using (2.1) and (2.2), we see that the probability that $\{\mathbf{V}_r \in \times_{i=m+1}^n B_i\}$ given $\mathbf{V}_{r-1}$ is given by

$$\begin{aligned}
\text{Prob}\{\mathbf{V}_r \in \times_{m+1}^n B_i | \mathbf{V}_{r-1}\} &= \int \prod_{i=m+1}^n P_{C_i}(B_i \cap C_i) \mathcal{D}_\gamma(dP) \\
&\geq \int \prod_{i=m+1}^n P(B_i \cap C_i) \mathcal{D}_\gamma(dP) \\
&\geq \frac{\prod_{i=m+1}^n \gamma(B_i \cap C_i)}{\prod_{i=m+1}^n [\alpha(R) + i - 1]} \\
&\geq \theta \prod_{i=m+1}^n \alpha_{C_i}(B_i)
\end{aligned}$$

where $\gamma = \alpha' + \sum_{i=m+1}^n \delta_{\mathbf{V}_{(r-1),i}}$ and $\theta = \frac{\prod_{m+1}^n \alpha(C_i)}{\prod_{i=m+1}^n [\alpha(R)+i-1]} > 0$ is a quantity independent of $B_1, \ldots, B_n$. This verifies conditions (4.3) and (4.4) with $n_0 = 1$ for the chain $\{\mathbf{V}_r\}$. Thus, (4.5) is also satisfied, and the conclusions of Theorems 1 and 2 hold. The distribution of $(P_r, \mathbf{V}_r)$ is a continuous function of the distribution of $\mathbf{V}_{r-1}$ not depending on r and can be written down as follows by using (2.1) and (2.2) once again:

$$\text{Prob}\{P_r \in E, \mathbf{V}_r \in \times_{m+1}^n B_i | \mathbf{V}_{r-1}\} = \int_{P \in E} \prod_{i=m+1}^n P_{C_i}(B_i) \mathcal{D}_\gamma(dP)$$

where $\gamma = \alpha' + \sum_{i=m+1}^n \delta_{\mathbf{V}_{(r-1)}^{(i)}}$. The convergence of the distributions of $\{\mathbf{V}_r\}$ together with the above remark implies the convergence in distribution of the whole Markov chain $(P_0, \mathbf{V}_0), (P_1, \mathbf{V}_1), (P_2, \mathbf{V}_2), \ldots$ follows.

References

Athreya, K. B., Doss, H., and Sethuraman, J. (1996). On the Convergence of the Markov Chain Simulation Method *To appear in Annals of Statistics*

Athreya, K. B. and Ney, P. (1978). A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.* **245** 493–501.

Doss, H. (1996) Bayesian Nonparametric Estimation of Incomplete Data via Successive Substitution Sampling *To appear in the Annals of Statistics.*

Ferguson, T. S. (1973) A Bayesian Analysis of Some Nonparametric Problems *Annals of Statistics* **1** 209–230.

Hoel, P., Port, S., and Stone, C. (1972). *Introduction to Stochastic Processes.* Houghton Mifflin, Boston.

Nummelin, E. (1984). *General Irreducible Markov Chains and Non-Negative Operators.* Cambridge University Press, Cambridge.

Orey, S. (1971). *Limit Theorems for Markov Chains Transition Probabilities.* Van Nostrand, New York.

Rao, C. R. (1965) *Linear Statistical Inference and its Applications*, John Wiley and Sons, New York.

Sethuraman, J. (1994) A Constructive Definition of Dirichlet Priors *Statistics Sinica* **4** 639–650.

SF 298 MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
		Technical Report

4. TITLE AND SUBTITLE

Title on Technical Report

5. FUNDING NUMBERS

DAAH04-93-G-0201

6. AUTHOR(S)

Author(s) listed on Technical Report

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)

Florida State Univ
Tallahassee, Florida 32306

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

ARO 31029.3-MA

11. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Abstract on Technical Report

14. SUBJECT TERMS

15. NUMBER IF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18
298-102